DOCUMENT RESUME

ED 166 224 TM 008 174

AUTHOR Lindley, Dennis V.
TITLE A Bayesian Solution for Two-Way Analysis of Variance. ACT Technical Bulletin No. 8.
INSTITUTION American Coll. Testing Program, Iowa City, Iowa. Research and Development Div.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Sep 72
GRANT OEG-0-72-0711
NOTE 36p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Bayesian Statistics; *Mathematical Models; Tests of Significance

ABSTRACT

The standard statistical analysis of data classified in two ways (say into rows and columns) is through an analysis of variance that splits the total variation of the data into the main effect of rows, the main effect of columns, and the interaction between rows and columns. This paper presents an alternative Bayesian analysis of the same situation that is appropriate for certain types of prior knowledge. It leads to a rather different treatment of the three factors just mentioned. This analysis proposes a four stage model. The first stage describes the dependence of the x's on the thetas; the second, that of the thetas on the alphas and betas; the third describes the structure of the alphas and betas; and a fourth stage is necessary to describe the prior distributions of mus. Appendices provide derivations and a numerical example. (Author/CTM)

ACT TECHNICAL BULLETIN NO. 8

A BAYESIAN SOLUTION FOR TWO-WAY ANALYSIS OF VARIANCE

by

Dennis V. Lindley

University College London

The Research and Development Division

The American College Testing Program

P. O. Box 168, Iowa City, Iowa 52240

September, 1972

A BAYESIAN SOLUTION FOR TWO-WAY ANALYSIS OF VARIANCE[1]

by

Dennis V. Lindley

University College London

SUMMARY

The standard statistical analysis of data classified in two ways (say into rows and columns) is through an analysis of variance that splits the total variation of the data into the main effect of rows, the main effect of columns, and the interaction between rows and columns. This paper presents an alternative Bayesian analysis of the same situation that is appropriate for certain types of prior knowledge. It leads to a rather different treatment of the three factors just mentioned.

[1]The research reported herein was performed pursuant to Grant No. OEG-0-72-0711 with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

In this paper, we consider the analysis of data (x_{ijk}) having the following probability structure. For given parameter values (θ_{ij}) and (σ_{ij}^2), the random variables x_{ijk} are independent and normally distributed with $E(x_{ijk}) = \theta_{ij}$ and $var(x_{ijk}) = \sigma_{ij}^2$: here $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$; and $k = 1, 2, \ldots, r_{ij}$.

An example where this model for data might be appropriate is where x_{ijk} is the performance of a subject in an educational test, the subject having been to School i and College j, there being r_{ij} such subjects and the suffix k serving to enumerate them. There, θ_{ij} would correspond to the true score of subjects from School i and College j on the test, and σ_{ij}^2 would measure their variability. Any analysis of the data would investigate what effects the school and college attended had on performance. At first, we shall confine attention to the case where the variabilities σ_{ij}^2 are all the same, equal to σ^2; and there are the same numbers of subjects in each group, so that $r_{ij} = r$, say. This is usually referred to as the orthogonal case, and its analysis is rather simpler than that for the general situation which is discussed toward the end of the paper. Rather than refer to schools and colleges, we shall use the neutral terms "rows" and "columns"; x_{ijk} is then the k^{th} observation in Row i and Column j, the data being conveniently laid out on the page in such a row and column formation.

Let us first recall how such data are traditionally analyzed. Any good textbook on statistics that deals with the two-way analysis of variance, with interaction, will provide details beyond the summary which follows: for example, Snedecor (1956, Chapter 11). We use the familiar "dot" notation for averages. Thus, $x_{i..} = \sum_{j,k} x_{ijk}/nr$, the mean of the data in Row i, the dots replacing the suffixes j and k over which summation has taken place. The usual analysis breaks up the total sum of squares about the overall mean, $\sum_{i,j,k} (x_{ijk} - x_{...})^2$, into at least four components. Firstly, there is the main effect of rows

$$nr\sum_i (x_{i..} - x_{...})^2 , \tag{1}$$

and secondly, that of columns

$$mr\sum_j (x_{.j.} - x_{...})^2 . \tag{2}$$

The third is the interaction between rows and columns

$$r\sum_{i,j} (x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2 , \tag{3}$$

and the last is the residual, or within groups, sum of squares

$$\sum_{i,j,k} (x_{ijk} - x_{ij.})^2 . \tag{4}$$

On division by their appropriate degrees of freedom, each of the first three may be tested against the last using the familiar F-test. If, for example, only the first test is significant, then the column and interaction effects are supposed zero and θ_{ij}, for all j, is estimated by $x_{i..}$. Comparisons between these means are effected by multiple-comparison procedures of which Scheffé's is, perhaps, the most popular.

This analysis, apart from being open to the usual criticisms that can be leveled against significance tests, is unsatisfactory in that it forces one into the position of having to be dogmatic about whether a particular effect exists, or not. Thus, several estimates of θ_{ij} are available depending on the results of the tests. Two are $x_{i..}$ (mentioned above) and $x_{i..} + x_{.j.} - x_{...}$ (if row and column, but no interaction, effects exist). A better procedure would be to estimate the size of each of these effects and estimate θ_{ij} accordingly. The methods developed below do just this and, for example, weight the row in which θ_{ij} appears heavily if the row effect appears to be large. Significance tests are, thereby, avoided.

For the one-way classification, where $E(x_{ik}) = \theta_i$, such an analysis has been given by Lindley (1971) and extended to other situations in the context

of a general theory by Lindley and Smith (1972). In this paper, we apply the results of the latter reference to obtain an estimate of θ_{ij} that uses $x_{ij.}$, $x_{i..}$, $x_{.j.}$, and $x_{...}$ in a balance that depends on the relative sizes of the main effects and interaction. In order to utilize this theory, it is necessary to describe the prior probability distribution of the (θ_{ij}) (and also σ^2, but in the first analysis this will be supposed known). In the one-way case, it was suggested that the joint distribution might reasonably have the property of exchangeability; that is, be invariant under any permutation of the suffixes. This property is clearly inappropriate in the two-way case as is seen by considering the joint distribution of a pair, θ_{ij} and θ_{rs}. Under exchangeability, this distribution is the same for any pair of (different) θ's, whereas it would be reasonable for the relation between θ_{ij} and θ_{is} $(j \neq s)$ in the same row to be different from that between θ_{ij} and θ_{rs} $(i \neq r)$ in different rows (and columns). In our example, knowledge of the performance of subjects at School i and College j might affect knowledge of subjects from the <u>same</u> school at another college, whereas it might say little about those from a <u>different</u> school at the college. We, therefore, have to express the prior ideas other than through exchangeability. We use, instead, a modified form of it.

Our prior opinions might lead us to think that the value of θ_{ij} is influenced both by the row and the column that it is in. If these effects are assumed additive, we might suppose

$$\theta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

where μ is an overall mean, (α_i) and (β_j) respectively describe row and column effects, and (γ_{ij}) represent independent error terms, say, normal with zero mean and variance σ_c^2. Alternatively expressed, we could say; given μ, (α_i), (β_j), and σ_c^2, the θ's are independent and normally distributed with

$$E(\theta_{ij}) = \mu + \alpha_i + \beta_j \tag{5}$$

and variance σ_c^2. The rows and columns might reasonably be exchangeable; and hence, given μ_a, μ_b, σ_a^2, σ_b^2, we might suppose the α's and β's independent and normally distributed with $E(\alpha_i) = \mu_a$, $E(\beta_j) = \mu_b$, $\mathrm{var}(\alpha_i) = \sigma_a^2$, and $\mathrm{var}(\beta_j) = \sigma_b^2$.

This model fits conveniently into the framework developed by Lindley and Smith. In their terminology, it is a four-stage model; the first stage describes the dependence of the x's on the θ's; the second, that of the θ's on the α's and β's; the third describes the structure of the α's and β's; and a fourth stage is necessary to describe the prior distributions of μ_a and μ_b. As in earlier examples, this distribution can be supposed diffuse and the variances for μ_a and μ_b allowed to tend to infinity. It is possible to proceed with the analysis of the four-stage form, but it is convenient to reduce it first to a three-stage version with a diffuse prior at the third and final stage: the two analyses are equivalent, except for one point to be discussed later in considering the variance estimation.

To derive the three-stage model, consider the distribution of the θ's, given μ, but not the α's and β's. From (5), it is clear that the covariances are given by

$$\mathrm{cov}(\theta_{ij}, \theta_{rs}) = 0, \quad i \neq r,\ j \neq s \ , \tag{6a}$$

$$\mathrm{cov}(\theta_{ij}, \theta_{is}) = \sigma_a^2, \quad j \neq s \ , \tag{6b}$$

$$\mathrm{cov}(\theta_{ij}, \theta_{rj}) = \sigma_b^2, \quad i \neq r \ , \tag{6c}$$

and

$$\mathrm{cov}(\theta_{ij}, \theta_{ij}) = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 \ . \tag{6d}$$

(The last is just the variance of θ_{ij}.) For example, the difference between (6a) and (6b) is just the distinction we were discussing above concerning

subjects from the same School (row) i . Consequently, a second stage, which replaces the second and third stages of the first model, supposes (θ_{ij}) has a multivariate normal distribution with covariance structure given by equation (6) and constant mean μ (now incorporating μ_a and μ_b). The third (and final) stage says the knowledge of μ is diffuse.

This is the model we suggest might be appropriate for some two-way analyses. We must emphasize that there may well exist two-way situations in which the above prior specification (in the second and third stages) is quite unsuitable. Before performing an analysis of the type suggested below, it must first be checked that the model is reasonably suitable. Our second- and third-stage forms are assumptions that may not always be realistic. For example, suppose the rows (schools) were of two types, say urban and rural, then the α's (in the four-stage form) would not be exchangeable for all i--perhaps, only within-urban and within-rural schools.

With this caution, let us summarize the model:

<u>First stage</u>. Given (θ_{ij}), σ^2; the (x_{ijk}) are normal and independent with $E(x_{ijk}) = \theta_{ij}$ and variance σ^2.

<u>Second stage</u>. Given μ, σ_a^2, σ_b^2, σ_c^2; the (θ_{ij}) have a multivariate normal distribution with dispersion matrix, given by equations (6), and $E(\theta_{ij}) = \mu$.

<u>Third stage</u>. The prior knowledge of μ is diffuse.

Our first object is, for given σ^2, σ_a^2, σ_b^2, and σ_c^2, to find the posterior distribution of the (θ_{ij}). It is easy to see that it will be multivariate normal; the means will then provide estimates of the (θ_{ij}), and the dispersion matrix will enable standard errors to be attached to these estimates. We later relax the conditions on the knowledge of the four variances and show how they too may be estimated, merely providing revised estimates

and standard errors for the θ's. Finally, we discuss the more general first stage where $var(x_{ijk}) = \sigma^2_{ij}$; and the number, r_{ij}, of observations varies from cell to cell.

The algebraic derivation of the estimates θ^*_{ij} of θ_{ij} is given in Appendix 1. It is there shown that

$$\theta^*_{ij} = \frac{r/\sigma^2}{r/\sigma^2 + \sigma_c^{-2}} (x_{ij.} - x_{i..} - x_{.j.} + x_{...})$$
$$+ \frac{r/\sigma^2}{r/\sigma^2 + 1/(\sigma_c^2 + n\sigma_a^2)} (x_{i..} - x_{...})$$
$$+ \frac{r/\sigma^2}{r/\sigma^2 + 1/(\sigma_c^2 + m\sigma_b^2)} (x_{.j.} - x_{...}) + x_{...} \qquad (7)$$

This is the main result of this paper. The form of this estimate is interesting. It depends on four aspects of the data: $x_{ij.}$, the mean of the observations in cell (i, j); $x_{i..}$, and $x_{.j.}$, the corresponding row and column means; and $x_{...}$, the overall mean. It is a weighted combination of this last, $x_{i..} - x_{...}$, the effect of the row, $x_{.j.} - x_{...}$, the effect of the column, and $x_{ij.} - x_{i..} - x_{.j.} + x_{...}$, the interaction effect. The weights depend on the variances σ_a^2, σ_b^2, and σ_c^2 in addition to the residual variance (from the data) σ^2. Some special cases are interesting. Suppose $\sigma_c^2 = 0$ so that, equation (5), θ_{ij} is a linear combination of the row and column effects and no interaction exists. Then, the first term in (7) vanishes, there is no contribution from the data-interaction effect, and θ^*_{ij} uses only $x_{i..}$, $x_{.j.}$, and $x_{...}$. This is an extreme case corresponding to the assumed lack of an interaction as indicated in the usual approach by a non-significant F-test for the interaction. If, in addition to $\sigma_c^2 = 0$, $\sigma_a^2 = 0$, the second term in (7) also vanishes and only the column effect appears from the data.

If $\sigma_a^2 = 0$ without σ_c^2 vanishing, the first and second terms in (7) combine to give a multiple of $(x_{ij.} - x_{.j.})$. These results generalize in a natural way those of Lindley (1971) for the one-way case in which similar weighted combinations occurred. Later, we shall see how to estimate the four variances and, hence, the weights.

To obtain the posterior variances and covariance of these estimates, write the weights in (7) as

$$\left.\begin{aligned} w_c &= \frac{r/\sigma^2}{r/\sigma^2 + \sigma_c^{-2}}, \\ w_a &= \frac{r/\sigma^2}{r/\sigma^2 + 1/(\sigma_c^2 + n\sigma_a^2)}, \\ w_b &= \frac{r/\sigma^2}{r/\sigma^2 + 1/(\sigma_c^2 + m\sigma_b^2)} \end{aligned}\right\} \qquad (8)$$

Then, (7) becomes

$$\theta_{ij}^* = w_c(x_{ij.} - x_{i..} - x_{.j.} + x_{...}) + w_a(x_{i..} - x_{...}) + w_b(x_{.j.} - x_{...}) + x_{...}$$

If we further put

$$nW_a = w_a - w_c, \quad mW_b = w_b - w_c, \quad mnW = w_c - w_a - w_b + 1 \qquad (9)$$

and put $w_c = W_c$, for symmetry, (7) can be written

$$\theta_{ij}^* = W_c x_{ij.} + nW_a x_{i..} + mW_b x_{.j.} + mnWx_{...} \qquad (10)$$

For reasons given in Appendix 1, the dispersion matrix for θ_{ij} is given by [compare equations (6)]

$$\mathrm{cov}(\theta_{ij}, \theta_{rs}) = W\sigma^2/r, \qquad i \neq r,\ j \neq s \ , \tag{11a}$$

$$\mathrm{cov}(\theta_{ij}, \theta_{is}) = (W_a + W)\sigma^2/r, \qquad j \neq s \ , \tag{11b}$$

$$\mathrm{cov}(\theta_{ij}, \theta_{rj}) = (W_b + W)\sigma^2/r \ , \qquad i \neq r \ , \tag{11c}$$

and

$$\mathrm{cov}(\theta_{ij}, \theta_{ij}) = (W_a + W_b + W_c + W)\sigma^2/r \ . \tag{11d}$$

These expressions are somewhat cumbersome since the W's are fairly complicated, but some results are a little easier. For example, consider the posterior variance of $\theta_{ij} - \theta_{is}$ $(j \neq s)$, that is the difference between Columns j and s in the same Row i . It is 2 $\mathrm{var}(\theta_{ij}) - 2\ \mathrm{cov}(\theta_{ij}, \theta_{is})$, which, from (11b) and (11d) is $(W_b + W_c)2\sigma^2/r$. For the means of rows (or columns), the results are easier still. For example, the variance of $\theta_{i\cdot} - \theta_{r\cdot}$, the difference between two rows (schools) averaged over columns (colleges) is $(i \neq r,\ j \neq s)$

$$\begin{aligned}
\mathrm{var}(\theta_{i\cdot} - \theta_{r\cdot}) &= n^{-2}\mathrm{var}(\textstyle\sum_j \theta_{ij} - \sum_s \theta_{rs}) \\
&= n^{-2}\Big[2n\ \mathrm{var}(\theta_{ij}) - 2n\ \mathrm{cov}(\theta_{ij}, \theta_{rj}) \\
&\qquad + 2n(n-1)\mathrm{cov}(\theta_{ij}, \theta_{is}) - 2n(n-1)\mathrm{cov}(\theta_{ij}, \theta_{rs})\Big] \\
&= \frac{2\sigma^2}{rn}\Big[W_a+W_b+W_c+W - (W_b+W) + (n-1)(W_a+W) - (n-1)W\Big]
\end{aligned}$$

from (11), and using (9), this is finally equal to

$$\frac{2\sigma^2}{rn} w_a = \frac{2}{n}\left(\frac{r}{\sigma^2} + \frac{1}{\sigma_c^2 + n\sigma_a^2}\right)^{-1} \ . \tag{12}$$

Since θ^*_{ij} [equation (7)] is the posterior mean, the mean of $\theta_{i.}$ is $\theta^*_{i.}$, which, from (7), is easily seen to be

$$w_a(x_{i..} - x_{...}) + x_{...} = w_a x_{i..} + (1 - w_a)x_{...} ,$$

a weighted average of $x_{i..}$ and $x_{...}$. Had $x_{i..}$ been used as an estimate, as standard theory would suggest, then the variance for $\theta_{i.} - \theta_{r.}$ quoted would be $2\sigma^2/rn$ rather than this times w_a, given by (12). Hence, our estimate is pulled toward the overall mean and has smaller variances when compared with other values. It follows that the usual multiple comparison procedures, such as Scheffé's, are unnecessary in our approach. The shift toward the mean and the reduced standard errors perform exactly the function that these orthodox procedures are designed to provide.

These estimates (and standard errors) depend upon knowledge of the four variances σ^2, σ^2_a, σ^2_b, and σ^2_c. In any application, these are typically unknown but can be estimated from the data. This is obvious for σ^2 but is also true for the others since there is replication of rows and columns. We proceed to discuss their estimation.

Lindley and Smith, in discussing the general theory, show that if we are content with posterior modes for estimates (rather than posterior means), we can continue to estimate θ_{ij} by equations (7) provided we insert, for the four variances, modal estimates of them. It will, therefore, suffice to find the posterior modes for the variances. It is inconvenient to do this within the context of the three-stage model because the compression of two stages into one results in σ^2_c (for the original second stage) being combined with σ^2_a and σ^2_b (from the third stage) in expressions like $\sigma^2_c + n\sigma^2_a$, and we have the difficulties familiar in components of variance problems (or what

is sometimes called Type II analysis of variance) of having to estimate σ_c^2 and $\sigma_c^2 + n\sigma_a^2$ separately, and hence σ_a^2 by subtraction, so leading to the possibility of negative estimates for σ_a^2, or even within the Bayesian framework, to difficult calculations. This can be avoided by using the four-stage model, when the procedure is essentially to estimate (α_i) and (β_j) and, hence, σ_a^2 by a multiple of $\sum_i (\alpha_i^* - \alpha_.^*)^2$; similarly, σ_b^2. Also, σ_c^2 can be found from the sums of squares of $\theta_{ij}^* - \alpha_i^* - \beta_j^*$ [see equation (5)]. Finally, σ^2 can be found, although the usual within sum of squares is not enough since θ_{ij} is, within the present theory, not estimated by $x_{ij.}$ as is usual. Hence, the within-sum underestimates the total variation that contributes to σ^2. All these ideas are straightforward generalizations of ideas contained in the papers to which reference has already been made.

The details of the calculation of the posterior modes are given in Appendix 2. Equations (2.3) and (2.4) provide estimates of (α_i) and (β_j), respectively. Notice that only the deviation from the mean is estimated, which is all that is necessary. Distinction should be made between the estimate of, for example, α_i by [equation (2.3)]

$$(\alpha_i - \alpha_.)^* = \frac{rn\sigma_a^2}{rn\sigma_a^2 + r\sigma_c^2 + \sigma^2}(x_{i..} - x_{...})$$

and that of $\theta_{i.}$ by

$$(\theta_{i.} - \theta_{..})^* = \frac{rn\sigma_a^2 + r\sigma_c^2}{rn\sigma_a^2 + r\sigma_c^2 + \sigma^2}(x_{i..} - x_{..})$$

[from (1.18), or (7) on summing over j, and a little simplification.] The difference is that $\theta_{i.}$ is the average for Row i over the columns used in the experiment, whereas α_i is a similar average not confined to the columns of the experiment. In particular, α_i^* is shrunk more toward the overall mean

than is $\theta^*_{i.}$, since the coefficient of the deviation $(x_{i..} - x_{...})$ is smaller in the former.

Equations (2.5) provide the estimates of the variances, using the estimates for (α_i) and (β_j) just obtained as well as those for (θ_{ij}) already calculated. Those estimates, in turn, depend on the variances, and so some iterative procedure has to be used. We suggest the following: Obtain initial estimates of the four variances from the usual analysis of variance expressions, expressions (1) to (4), divided by their respective degrees of freedom. These will be unsatisfactory estimates but will serve to provide weights to be used to estimate the θ's [equation (7)] and the α's and β's. With these estimated, new values for the variances can be found from equations (2.5) and the cycle repeated until convergence.

Notice that the estimates (2.5) involve quantities derived from the prior distributions of the variances. There is no objection to putting ν, corresponding to σ^2, equal to zero; but the remaining values ν_a, ν_b, ν_c cannot be ignored. The difficulty is that if σ^2_a, σ^2_b, or σ^2_c are small in comparison with σ^2 (or more correctly σ^2/r), there is little information in the data from which to estimate them since the variation in the $(x_{ij.})$ is mostly due to σ^2. In this case, the prior knowledge is clearly important and so naturally arises in any estimation procedure. If σ^2_a, for example, is large; its estimation is easier, and in (2.5b), the sum of squares for α^*_i will dominate $\nu_a\lambda_a$ unless the latter is large: that term and ν in the denominator may be ignored.

Whilst the estimates for θ_{ij}, given the variances, are almost certainly satisfactory; it may be possible to improve the estimation of the variances in comparison with the methods given in this paper; and we hope to study the problem in more detail later. In the meantime, it might be reasonable to

guess that the term mn in the denominator of (2.5d) might be replaced by the degrees of freedom (m - 1)(n - 1). In deriving modes, rather than means, the usual integrations that remove degrees of freedom do not take place, and hence, the divisor always involves the total number, here mn, of parameters. Another way of looking at it is to appreciate that the modes of marginal distributions are not the components of the modes of the whole distribution.

The discussion has so far been confined to the case where there is the same number, r, of observation in each cell. Suppose now that there are r_{ij} observations in the cell in the i^{th} row and j^{th} column. In this case, it is not possible to obtain simple expressions for the estimates θ^*_{ij} as in equation (7). Instead, we have to be content with linear equations for them which can then be solved numerically in any particular case. The estimates of (α_i) and (β_j) follow with minor modifications as do the estimation of the variances. Details are given in Appendix 3.

The last generalization we make is to the case where the within-cell variance σ^2_{ij} is not constant. In most applications, r_{ij} will not be large enough to effect a good estimation of σ^2_{ij}; but if the latter are assumed connected in some way, then sensible estimation may be possible. We have been able to make progress in the case where all the (σ^2_{ij}) are exchangeable. Ideally perhaps, one could make a modified exchangeability assumption as we have with the means, but I have not been able to develop a satisfactory procedure. Details with the full exchangeability assumption are given in Appendix 4. Appendix 5 summarizes the calculations required in the general case. Finally, Appendix 6 provides a simple numerical example.

APPENDIX 1: Posterior Distribution of the Cell Means Assuming the Variances Known

When writing out vectors of elements depending on two or more suffixes, we shall use a lexicographical order: thus,

$$\theta_1^T = (\theta_{11}, \theta_{12}, \ldots, \theta_{1n}, \theta_{21}, \theta_{22}, \ldots, \theta_{mn}) .$$

The three-stage model is exactly in the linear framework developed by Lindley and Smith, and their corollary 2 [equations (16) and (17)] shows that the posterior distribution of (θ_{ij}) is normal with first and second moments there stated. Their notation is

First stage. $E(x) = A_1\theta_1$, dispersion matrix C_1 .

Second stage. $E(\theta_1) = A_2\mu$, dispersion matrix C_2 .

Then, the posterior distribution of θ_1 is $N(Dd, D)$ with

$$D^{-1} = A_1^T C_1^{-1} A_1 + C_2^{-1} - C_2^{-1} A_2 (A_2^T C_2^{-1} A_2)^{-1} A_2^T C_2^{-1} \tag{1.1}$$

and

$$d = A_1^T C_1^{-1} x . \tag{1.2}$$

We proceed to evaluate (1.1) and (1.2). The matrix C_2 is given in equations (6): thus, the element in the row corresponding to θ_{ij} and column corresponding to θ_{is} $(j \neq s)$ is σ_a^2, and others similarly. The inversion required for (1.1) is most easily accomplished by solving the equations in z, $C_2 z = a$. Written out in full, these are

$$\sigma_c^2 z_{ij} + n\sigma_a^2 z_{i\cdot} + m\sigma_b^2 z_{\cdot j} = a_{ij} , \tag{1.3}$$

using the "dot" notation. Summing over i and j, we have

$$(\sigma_c^2 + n\sigma_a^2 + m\sigma_b^2)z_{..} = a_{..}$$

or

$$z_{..} = a_{..}/v_{mn} \tag{1.4}$$

where

$$v_{mn} = \sigma_c^2 + n\sigma_a^2 + m\sigma_b^2 \quad . \tag{1.5}$$

Summing (1.3) over j, we similarly obtain

$$(\sigma_c^2 + n\sigma_a^2)z_{i.} + m\sigma_b^2 z_{..} = a_{i.}$$

which, on using (1.4), can be written

$$z_{i.} = (a_{i.} - m\sigma_b^2 a_{..}/v_{mn})/v_n \tag{1.6}$$

where

$$v_n = \sigma_c^2 + n\sigma_a^2 \quad . \tag{1.7}$$

Similarly,

$$z_{.j} = (a_{.j} - n\sigma_a^2 a_{..}/v_{mn})/v_m \tag{1.8}$$

where

$$v_m = \sigma_c^2 + m\sigma_b^2 \quad . \tag{1.9}$$

Substitution of (1.6) and (1.7) into (1.3) gives

$$z_{ij} = \sigma_c^{-2}\left[a_{ij} - \frac{n\sigma_a^2}{v_n}(a_{i.} - m\sigma_b^2 a_{..}/v_{mn}) - \frac{m\sigma_b^2}{v_m}(a_{.j} - n\sigma_a^2 a_{..}/v_{mn})\right] \quad . \tag{1.10}$$

Since $\underset{\sim}{z} = \underset{\sim}{C}_2^{-1}\underset{\sim}{a}$, identification of terms on the right-hand shows that $\underset{\sim}{C}_2^{-1}$ has the same structure as $\underset{\sim}{C}_2$ itself [equations (6)]. For example, all the terms in rows (i, j) and columns (r, s) with $i \neq r$, $j \neq s$ are the same. From (1.10), the terms are

$$i \neq r,\ j \neq s: \quad h\ , \tag{1.11a}$$

$$i = r,\ j \neq s: \quad f + h\ , \tag{1.11b}$$

$$i \neq r,\ j = s: \quad g + h\ , \tag{1.11c}$$

$$i = r,\ j = s: \quad e + f + g + h\ , \tag{1.11d}$$

where h is the coefficient of $mna_{..}$ in (1.10). That is,

$$h = \frac{\sigma_a^2\sigma_b^2}{\sigma_c^2 v_{mn}}\left(\frac{1}{v_n} + \frac{1}{v_m}\right)\ , \tag{1.12a}$$

f is the coefficient of $na_{i.}$ in (1.10); namely,

$$f = -\ \sigma_a^2/\sigma_c^2 v_n\ . \tag{1.12b}$$

Similarly, g is the coefficient of $ma_{.j}$, so

$$g = -\ \sigma_b^2/\sigma_c^2 v_m\ , \tag{1.12c}$$

and e corresponds to a_{ij}; namely,

$$e = \sigma_c^{-2}\ . \tag{1.12d}$$

We note for future reference that summation of (1.10) over i and j gives $z_{..} = a_{..}(e + nf + mg + mnh)$ so that, on comparison with (1.4),

$$e + nf + mg + mnh = v_{mn}^{-1}\ . \tag{1.13}$$

Having evaluated $\underset{\sim}{C}_2^{-1}$, ~~we~~ now return to (1.1). $\underset{\sim}{A}_2$ is easily seen to be a vector, all of whose elements are unity. Hence, $\underset{\sim}{A}_2^T\underset{\sim}{C}_2^{-1}$ is a (row) vector, all of whose elements are $e + nf + mg + mnh = v_{mn}^{-1}$ [from (1.13)]. Hence, $\underset{\sim}{A}_2^T\underset{\sim}{C}_2^{-1}\underset{\sim}{A}_2 = mnv_{mn}^{-1}$. Simple calculation shows that $\underset{\sim}{C}_2^{-1}\underset{\sim}{A}_2(\underset{\sim}{A}_2^T\underset{\sim}{C}_2^{-1}\underset{\sim}{A}_2)^{-1}\underset{\sim}{A}_2^T\underset{\sim}{C}_2^{-1}$ is a matrix, every element of which is $(mnv_{mn})^{-1}$.

Reference to the first stage of the model shows easily that $\underset{\sim}{A}_1^T \underset{\sim}{C}_1^{-1} \underset{\sim}{A}_1$ is a diagonal matrix with every diagonal element equal to r/σ^2. Consequently, $\underset{\sim}{D}^{-1}$ [equation (1.1)] is a matrix of the same form as $\underset{\sim}{C}_2^{-1}$ [equations (1.11)]; but with e replaced by $e + r/\sigma^2 = e'$, say, and h by $h - (mnv_{mn})^{-1} = h'$, say. The values of f and g are unaltered. Further consideration of the first stage of the model shows that $\underset{\sim}{d}$ is a vector whose (i, j) element is $x_{ij.} r/\sigma^2$.

If θ^*_{ij} denotes the estimate of θ_{ij}, that is, the posterior mean of their joint distribution; the corollary quoted about shows that $\underset{\sim}{\theta}^* = \underset{\sim}{D}\underset{\sim}{d}$, or $\underset{\sim}{D}^{-1}\underset{\sim}{\theta}^* = \underset{\sim}{d}$. Inserting the values of $\underset{\sim}{D}^{-1}$ and $\underset{\sim}{d}$ just obtained and writing these equations out in full, we have

$$e'\theta^*_{ij} + nf\theta^*_{i.} + mg\theta^*_{.j} + mnh'\theta^*_{..} = x_{ij.} r/\sigma^2 \tag{1.14}$$

[compare equations (1.3)]. These equations are most easily solved by writing

$$\left.\begin{aligned} \phi^*_{ij} &= \theta^*_{ij} - \theta^*_{i.} - \theta^*_{.j} + \theta^*_{..} , \\ \phi^*_{i.} &= \theta^*_{i.} - \theta^*_{..} , \\ \phi^*_{.j} &= \theta^*_{.j} - \theta^*_{..} , \end{aligned}\right\} \tag{1.15}$$

and

$$\left.\begin{aligned} y_{ij} &= x_{ij.} - x_{i..} - x_{.j.} + x_{...} , \\ y_{i.} &= x_{i..} - x_{...} , \\ y_{.j} &= x_{.j.} - x_{...} . \end{aligned}\right\} \tag{1.16}$$

We can then rewrite (1.14) as

$$e'\phi^*_{ij} + (e' + nf)\phi^*_{i.} + (e' + mg)\phi^*_{.j} + (e' + nf + mg + mnh')\theta^*_{..}$$

$$= (y_{ij} + y_{i.} + y_{.j} + x_{..})r/\sigma^2 . \tag{1.17}$$

We note, from (1.13), and the fact that $e' = e + r/\sigma^2$, $h' = h - (mnv_{mn})^{-1}$; that $e' + nf + mg + mnh' = r/\sigma^2$.

Summation of (1.17) over i and j gives $\theta^*_{..} = x_{..}$, over j alone gives $(e' + nf)\phi^*_{i.} = y_{i.} r/\sigma^2$ or

$$\phi^*_{i.} = \frac{r/\sigma^2}{r/\sigma^2 + v_n^{-1}} \; y_{i.} \tag{1.18}$$

on inserting the values for $e' = e + r/\sigma^2$, e, (1.12d) and f, (1.12b).

Similarly,

$$\phi^*_{.j} = \frac{r/\sigma^2}{r/\sigma^2 + v_m^{-1}} \; y_{.j} \quad , \tag{1.19}$$

and inserting these values into (1.17),

$$\phi^*_{ij} = \frac{r/\sigma^2}{r/\sigma^2 + \sigma_c^{-2}} \; y_{ij} \quad . \tag{1.20}$$

Returning to the original form in terms of θ^*_{ij} and x_{ij}, we easily obtain the expressions given in (7).

The dispersion matrix for these estimates (that is, the dispersion matrix of the posterior normal distribution) is, by the corollary, $\underset{\sim}{D}$. The equations just solved are $\theta^* = \underset{\sim}{D}\underset{\sim}{d}$, so $\underset{\sim}{D}$ may be found by taking the coefficients of the elements, $x_{ij.} r/\sigma^2$, of $\underset{\sim}{d}$ in the solutions. For example, to obtain the covariance of θ^*_{ij} and θ^*_{rs} with $i \neq r$, $j \neq s$, it is only necessary to take the coefficient of $x_{rs.} r/\sigma^2$ in the expression for θ^*_{ij}. In the notation given by (8) and (9), this is easily to be seen from (10), W since $x_{rs.}$ only occurs in $x_{...}$, these with coefficient W. All the expressions given in equations (11) can be obtained in the same way.

APPENDIX 2: Estimation of the Variance Components

In the four-stage model, described by (5) and the following sentence, the joint probability distribution of all the random quantities (x_{ijk}), (θ_{ij}), (α_i), and (β_j), after integration with respect to the diffuse priors of μ, μ_a, and μ_b, is easily seen to be proportional to

$$\sigma^{-mnr}\sigma_c^{-mn+1}\sigma_a^{-m+1}\sigma_b^{-n+1}\exp\left[-\tfrac{1}{2}\sum_{i,j,k}(x_{ijk}-x_{ij.})^2\Big/\sigma^2\right]$$

$$\times\exp\left\{-\tfrac{1}{2}\left[\frac{r}{\sigma^2}\sum_{i,j}(x_{ij.}-\theta_{ij})^2+\frac{1}{\sigma_c^2}\sum_{i,j}(\theta_{ij}-\theta_{..}-\alpha_i+\alpha_.-\beta_j+\beta_.)^2\right.\right.$$

$$\left.\left.+\frac{1}{\sigma_a^2}\sum_i(\alpha_i-\alpha_.)^2+\frac{1}{\sigma_b^2}\sum_j(\beta_j-\beta_.)^2\right]\right\} . \tag{2.1}$$

There, the total sum of squares for the data has been broken into the two components within- and between-cells. Differentiation with respect to the θ's, α's, and β's, and equating the results to zero gives modal estimates for these parameters. It is not difficult to verify that for θ_{ij} is exactly θ^*_{ij} given by the three-stage model in equation (7). We proceed to find the corresponding modes (α^*_i) and (β^*_j). The result of differentiating (2.1) with respect to α_i is easily seen to be

$$\frac{n}{\sigma_c^2}\phi_{i.}-(\alpha_i-\alpha_.)\left(\frac{n}{\sigma_c^2}+\frac{1}{\sigma_a^2}\right) , \tag{2.2}$$

where $\phi_{i.}=\theta_{i.}-\theta_{..}$ [cf (1.15)]. Equating this to zero and using the estimate of $\phi^*_{i.}$ [equation (1.18)], we easily obtain

$$(\alpha_i-\alpha_.)^*=\frac{rn\sigma_a^2}{rn\sigma_a^2+r\sigma_c^2+\sigma^2}(x_{i..}-x_{...}) . \tag{2.3}$$

Similarly,

$$(\beta_j-\beta_.)^*=\frac{rm\sigma_b^2}{rm\sigma_b^2+r\sigma_c^2+\sigma^2}(x_{.j.}-x_{...}) . \tag{2.4}$$

With these estimates, it is an easy matter to obtain equations for the modal estimates of the four variance components σ^2, σ_a^2, σ_b^2, and σ_c^2. Suppose these have independent prior distributions which are all inverse $-\chi^2$. Specifically, let

$$\frac{\nu\lambda}{\sigma^2} \sim \chi_\nu^2 \quad \frac{\nu_t\lambda_t}{\sigma_t^2} \sim \chi_{\nu_t}^2 \quad (t = a, b, c) \quad .$$

Multiplication of the distribution (2.1), by this prior, gives the posterior distribution of all the parameters, including the variances, apart from constant factors. The modal equations for the variances are straightforward since the expression factors into four parts, each depending on one of the variances. The results are (we use s^2 for an estimate of σ^2 rather than the asterisk notation used with the other parameters)

$$s^2 = \left[\nu\lambda + S_w + r \sum_{i,j} (x_{ij.} - \theta_{ij}^*)^2\right] \Big/ (mnr + \nu + 2) \tag{2.5a}$$

$$s_a^2 = \left[\nu_a\lambda_a + \sum_i(\alpha_i^* - \alpha_.^*)^2\right] \Big/ (m + \nu_a + 1) \tag{2.5b}$$

$$s_b^2 = \left[\nu_b\lambda_b + \sum_j(\beta_j^* - \beta_.^*)^2\right] \Big/ (n + \nu_b + 1) \tag{2.5c}$$

$$s_c^2 = \left[\nu_c\lambda_c + \sum_{i,j} (\theta_{ij}^* - \theta_{..}^* - \alpha_i^* + \alpha_.^* - \beta_j^* + \beta_.^*)^2\right] \Big/ (mn + \nu_c + 1), \tag{2.5d}$$

where $S_w = \sum_{i,j,k} (x_{ijk} - x_{ij.})^2$, the usual within-cells sum of squares. For reasons given in the main text, mn in the denominator of (2.5d) can probably be replaced by $(m - 1)(n - 1)$.

APPENDIX 3: Unequal Numbers of Observations in the Cells

In this appendix, we consider the case where cell (i, j) contains r_{ij} observations, not all equal. Since the change from constant $r_{ij} = r$ only affects the first stage of the model, the calculations in Appendix 1, on the second stage, are unaffected. However, $\underset{\sim}{A}_1^T \underset{\sim}{C}_1^{-1} \underset{\sim}{A}_1$ will be a diagonal matrix with diagonal entries r_{ij}/σ^2. The result will be that $\underset{\sim}{D}^{-1}$ will not have a constant diagonal entry; and e in C_2^{-1} will be replaced, not by $e' = e + r/\sigma^2$, but by entries $e + r_{ij}/\sigma^2$. Equations (1.14) will, therefore, read

$$(e + r_{ij}/\sigma^2)\theta^*_{ij} + nf\theta^*_{i.} + mg\theta^*_{.j} + mnh'\theta^*_{..} = x_{ij.}r_{ij}/\sigma^2 . \qquad (3.1)$$

It does not seem possible to write down the solution to these at all simply and resort must be had to numerical calculation in any particular case. The matrix on the left-hand side of (3.1) is the inverse of the posterior dispersion matrix, $\underset{\sim}{D}$, and this too will have to be found numerically. It is not, therefore, possible to give formulae for the variances and covariances of θ^*_{ij}, generalizing equations (11).

With the θ's estimated, the argument leading to (2.2) is unaffected and the α's may be found from

$$(\alpha_i - \alpha_.)^* = \frac{n\sigma_a^2}{\sigma_c^2 + n\sigma_a^2} (\theta_{i.} - \theta_{..})^* . \qquad (3.2)$$

Similarly,

$$(\beta_j - \beta_.)^* = \frac{m\sigma_b^2}{\sigma_c^2 + m\sigma_b^2} (\theta_{.j} - \theta_{..})^* . \qquad (3.3)$$

Note that equations (2.3) and (2.4) are no longer available.

Finally, the estimation of σ_a^2, σ_b^2, and σ_c^2 [equations (2.5b-d)] is unaltered, but the new estimation of σ^2 is

$$s^2 = \left[\nu\lambda + S_w + \sum_{i,j} r_{ij}(x_{ij.} - \theta^*_{ij})^2\right] \Big/ (R + \nu + 2) \tag{3.4}$$

replacing (2.5a). There, $R = \sum_{i,j} r_{ij}$.

Notice the nonorthogonality problems that arise in the usual approach--for example, the nonindependence of sum of squares--does not matter here. Nevertheless, the complicated form of the posterior dispersion matrix does make it much more difficult to describe and understand the analysis; and for this reason, the balanced design is much to be preferred.

APPENDIX 4: Exchangeability and Variance Estimation

The main part of this appendix is virtually independent of the rest of the paper, but the results obtained therein are then applied to the two-way analysis when the within-cell variances σ^2_{ij} are not constant.

The estimation of variances has been discussed by Lindley (1971), but the analysis there contains some flaws; and we, therefore, approach the problem afresh. The simplest case of variance estimation is where there are m independent samples, each from a normal distribution of known, zero mean but unknown variance. Let the i^{th} sample have variance ϕ_i and denote the data sum of squares about the mean (i.e., zero) by S_i^2; this will have n_i degrees of freedom where n_i is the size of that sample. The (S_i^2) form a set of sufficient statistics, and the likelihood for the data is proportional to

$$\prod_{i=1}^{m} \left[\exp\left(-\tfrac{1}{2} \frac{S_i^2}{\phi_i} \right) \left(\frac{1}{\phi_i} \right)^{\frac{1}{2} n_i} \right] \qquad (4.1)$$

Suppose now that the prior opinions of the variances are that they are exchangeable. One way of achieving this is to suppose the (ϕ_i) themselves a random sample from some distribution: indeed, if the exchangeability is to hold for every m, then this is the only way to achieve it. It is convenient to suppose this distribution to be of the form conjugate to (4.1), namely, inverse $-\chi^2$. Specifically, we suppose the distribution of ϕ_i to be such that, for given ν and σ^2, $\nu\sigma/\phi_i$ is χ^2 with ν degrees of freedom. There, ν and σ^2 are hyperparameters, σ^2 being a measure of location for ϕ_i (and therefore, by the exchangeability, of every variance) and ν measuring the precision of that distribution. The prior distribution of the ϕ's, given the hyperparameters, is therefore,

$$\prod_{i=1}^{m}\left[\exp\left(-\tfrac{1}{2}\frac{\nu\sigma^2}{\phi_i}\right)\frac{(\nu\sigma^2)^{\frac{1}{2}\nu}}{\phi_i^{\frac{1}{2}\nu+1}}\frac{1}{2^{\frac{1}{2}\nu}(\frac{1}{2}\nu-1)!}\right].$$

This may be rewritten

$$\exp\left(-\tfrac{1}{2}\frac{m\nu\sigma^2}{H}\right)\left[\frac{(\nu\sigma^2)^{\frac{1}{2}\nu}}{G^{\frac{1}{2}\nu+1}}\frac{1}{2^{\frac{1}{2}\nu}(\frac{1}{2}\nu-1)!}\right]^m, \qquad (4.2)$$

where

$$\left.\begin{aligned}\Sigma\frac{1}{\phi_i} &= \frac{m}{H}\\ \text{and}\qquad \Pi\phi_i &= G^m ,\end{aligned}\right\} \qquad (4.3)$$

so that G and H are, respectively, the geometric and harmonic means of the variances we are trying to estimate.

The next stage is the assignment of a prior distribution for ν and σ^2. In the earlier paper, equation (16) of Lindley (1971), we assigned a distribution of σ^2, given ν; thus, making these two dependent. It seems more natural to think of them as independent since they measure quite different features of the distributions of the ϕ's. Suppose then that $\lambda\sigma^2$ is distributed as χ^2 on δ degrees of freedom, δ and λ being known constant values, independent of ν whose distribution will be discussed below. Since the mean of $\lambda\sigma^2$ is δ, $\delta\lambda^{-1}$ is our prior estimate of any ϕ_i. The value of δ reflects the precision attached to this estimate and would usually be small. The density of σ^2 is then proportional to

$$\exp\left[-\tfrac{1}{2}\lambda\sigma^2\right](\sigma^2)^{\frac{1}{2}\delta-1}. \qquad (4.4)$$

We have to multiply (4.2) by (4.4) and integrate the result with respect to σ^2. The only terms that involve σ^2 are

$$\exp\left[-\tfrac{1}{2}\left(\frac{m\nu}{H}+\lambda\right)\sigma^2\right](\sigma^2)^{\frac{1}{2}\nu m+\frac{1}{2}\delta-1} ,$$

and the integration gives

$$\frac{2^{\frac{1}{2}(\nu m+\delta)}[\tfrac{1}{2}(\nu m+\delta)-1]!}{\left(\frac{m\nu}{H}+\lambda\right)^{\frac{1}{2}(\nu m+\delta)}} .$$

Restoring the terms, so far omitted from (4.2), we get, apart from constants,

$$\frac{[\tfrac{1}{2}(\nu m+\delta)-1]!}{[(\tfrac{1}{2}\nu-1)!]^m}\;\frac{\nu^{\frac{1}{2}\nu m}}{G^{(\frac{1}{2}\nu+1)m}}\;\frac{1}{\left(\frac{m\nu}{H}+\lambda\right)^{\frac{1}{2}(\nu m+\delta)}} . \qquad (4.5)$$

This complicated expression can be simplified using Stirling's formula for the factorial function. Its most convenient form for our purpose is

$$\log(a\nu+b)! \sim c+(a\log a-a)\nu+a\nu\log\nu+(b+\tfrac{1}{2})\log\nu$$

for constants a, b, and c . The logarithm of (4.5) is then, apart from a constant which does not involve the ϕ's, and omitting terms of order ν^{-1}

$$\tfrac{1}{2}m\log(H/G).\nu+\tfrac{1}{2}(m-1)\log\nu-(m\log G+\tfrac{1}{2}\lambda H-\tfrac{1}{2}\delta\log H) .$$

Hence, (4.5) is, approximately, equal to

$$\exp\left(-\tfrac{1}{2}m\log\frac{G}{H}.\nu\right).\nu^{\frac{1}{2}(m-1)}e^{-\frac{1}{2}\lambda H}H^{\frac{1}{2}\delta}G^{-m} . \qquad (4.6)$$

Finally, suppose ν has a prior density proportional to

$$\exp(-\tfrac{1}{2}\lambda'\nu)\nu^{\frac{1}{2}\delta'-1} . \qquad (4.7)$$

The product of (4.6) and (4.7) is then easily integrated with respect to ν to give

$$\frac{e^{-\frac{1}{2}\lambda H}H^{\frac{1}{2}\delta}}{G^m\left(m\log\frac{G}{H}+\lambda'\right)^{\frac{1}{2}(m-1)+\frac{1}{2}\delta'}} . \qquad (4.8)$$

On multiplying by the likelihood, we have finally for the posterior distribution of the (ϕ_i) a value proportional to

$$\exp\left[-\tfrac{1}{2}\left(\Sigma \frac{S_i^2}{\phi_i} + \lambda H\right)\right] H^{\frac{1}{2}\delta}\left[\left(\prod_i \phi_i^{\frac{1}{2}n_i}\right) G^m\left(m \log \frac{G}{H} + \lambda'\right)^{\frac{1}{2}(m-1)+\frac{1}{2}\delta'}\right]^{-1} . \quad (4.9)$$

We proceed to find the modes of this distribution and to use these as estimates of the variances. Taking logarithms and differentiating (4.9), we have

$$\tfrac{1}{2}\frac{S_i^2}{\phi_i} - \tfrac{1}{2}\lambda \frac{H^2}{m\phi_i^2} + \tfrac{1}{2}\delta \frac{H}{m\phi_i^2} - \tfrac{1}{2}\frac{n_i}{\phi_i} - m\frac{1}{G}\frac{G}{m\phi_i}$$

$$- \left[\tfrac{1}{2}(m-1) + \tfrac{1}{2}\delta'\right] \frac{1}{\left(m \log \frac{G}{H} + \lambda'\right)} \left(\frac{1}{\phi_i} - \frac{H}{\phi_i^2}\right) = 0 .$$

(In obtaining this result, the derivatives

$$\frac{\partial H}{\partial \phi_i} = \frac{H^2}{m\phi_i^2} \quad \text{and} \quad \frac{\partial G}{\partial \phi_i} = \frac{G}{m\phi_i} ,$$

which are easily verified, have been used.) Consequently, the estimate ϕ_i^* of ϕ_i is given by

$$\phi_i^*\left[n_i + 2 + \frac{\frac{1}{2}(m-1) + \frac{1}{2}\delta'}{m \log \frac{G}{H} + \lambda'}\right] = \left[S_i^2 - \lambda H^2/m + \delta H/m + \frac{\frac{1}{2}(m-1) + \frac{1}{2}\delta'}{m \log \frac{G}{H} + \lambda'} H\right] . \quad (4.10)$$

This rather complicated expression for ϕ_i^* can be simplified. If we put $\lambda = \delta = 0$, we are [equation (4.4)] effectively assuming that we have little prior knowledge of σ^2, and we have the usual prior for a variance proportional to σ^{-2}. This causes no convergence problems in (4.10). We cannot do the same for ν [equation (4.7)], but $\delta' = 1$ will simplify things a little [for then,

$\frac{1}{2}(m-1) + \frac{1}{2}\delta' = \frac{1}{2}m]$, while avoiding convergence problems and yet representing diffuse knowledge of ν . (4.10) then becomes

$$\phi_i^* = \frac{(n_i + 2)s_i^2 + \frac{1}{\log\frac{G}{H} + \lambda''} H}{(n_i + 2) + \frac{1}{\log\frac{G}{H} + \lambda''}} \tag{4.11}$$

where $\lambda'' = \lambda'/m$ and $s_i^2 = S_i^2/(n_i + 2)$.

The form of (4.11) is informative. ϕ_i^* is a weighted average of the usual estimate, S_i^2 (apart from a divisor $n_i + 2$ instead of n_i) and the harmonic mean of the ϕ's. (In this mean, we can conveniently replace ϕ_i^* by s_i^2 .) Hence, we see that the estimates are pulled toward the harmonic mean just as the estimates of means move to the arithmetic mean. The weight attached to the mean is the reciprocal of $(\log\frac{G}{H} + \lambda'')$ and increases as the geometric and harmonic means become more disparate (note that $G \geq H$). It is not possible to let $\lambda'' = 0$, since then, infinite weight is attached to the mean value with $G = H$.

Now, let us apply these results to the two-way analysis of variance. In the four-stage model, the probability distribution will be as (2.1) except that the terms involving σ^2 will be replaced by

$$\prod_{ij}\sigma_{ij}^{-r_{ij}}\exp\left[-\tfrac{1}{2}\sum_{i,j,k}(x_{ijk} - x_{ij.})^2/\sigma_{ij}^2 - \tfrac{1}{2}\sum_{i,j}(x_{ij.} - \theta_{ij})^2 r_{ij}/\sigma_{ij}^2\right] . \tag{4.12}$$

On writing,

$$S_{ij}^2(\theta_{ij}) = \sum_k(x_{ijk} - x_{ij.})^2 + (x_{ij.} - \theta_{ij})^2 r_{ij} \quad , \tag{4.13}$$

this becomes

$$\prod_{i,j}\sigma_{ij}^{-r_{ij}}\exp\left[-\tfrac{1}{2}S_{ij}^2(\theta_{ij})/\sigma_{ij}^2\right] . \tag{4.14}$$

This is a likelihood of the same form as (4.1) with (bearing in mind the double suffixes) σ_{ij}^2 for ϕ_i, $S_{ij}^2(\theta_{ij})$ for S_i^2, r_{ij} for n_i, and mn for m .

We now suppose the σ_{ij}^2 to be exchangeable. This may not be appropriate because it fails to exploit the row and column structure of the layout; but as a first approximation, it might be reasonable. If we do this, the appropriate estimate of σ_{ij}^2 is given by the equivalent of (4.11): that is,

$$s_{ij}^2 = \frac{S_{ij}^2(\theta_{ij}^*) + \rho H}{r_{ij} + 2 + \rho} \qquad (4.15)$$

with

$$\rho^{-1} = \log \frac{G}{H} + \lambda'' \quad ,$$

G and H being, respectively, the geometric and harmonic means of the (s_{ij}^2).

APPENDIX 5: Methods of Calculation

In this appendix, we describe in summary form the steps to be carried out in calculating the estimates for the general case of unequal r_{ij} and σ^2_{ij}.

(1) Calculate from the data the basic statistics, $(x_{ij.})$ and $\left[\sum_k (x_{ijk} - x_{ij.})^2\right]$. Insert prior values for λ' (4.7)--λ'' (in 4.15) is λ'/mn--ν_t, λ_t $(t = a, b, c)$ [for equations (2.5b-d)].

(2) Calculate initial estimates of σ^2_a, σ^2_b, σ^2_c, and a pooled estimate s^2 of σ^2_{ij} using

$$s^2_a = rn\sum_i (x_{i..} - x_{...})^2/(m - 1) \quad ,$$

$$s^2_b = rm\sum_j (x_{.j.} - x_{...})^2/(n - 1) \quad ,$$

$$s^2_c = r\Sigma(x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2/(m - 1)(n - 1) \quad ,$$

and

$$s^2 = \sum_{i,j,k} (x_{ijk} - x_{ij.})^2 / \sum_{i,j} (r_{ij} - 1) \quad .$$

(3) With these estimates replacing σ^2_a, σ^2_b, σ^2_c, and σ^2, solve equations (3.1) for θ^*_{ij}. In these equations,

$$e = \sigma_c^{-2},\ nf = -n\sigma^2_a/\sigma^2_c(\sigma^2_c + n\sigma^2_a),\ mg = -m\sigma^2_b/\sigma^2_c(\sigma^2_c + m\sigma^2_b)$$

and

$$mnh' = \left[\frac{mn\sigma^2_a\sigma^2_b}{\sigma^2_c}\left(\frac{1}{\sigma^2_c + n\sigma^2_a} + \frac{1}{\sigma^2_c + m\sigma^2_b}\right) - 1\right] \Big/ (\sigma^2_c + n\sigma^2_a + m\sigma^2_b) \quad .$$

(4) Still using these estimates, find $(\alpha_i - \alpha_.)^*$ and $(\beta_j - \beta.)^*$ from equations (3.2) and (3.3).

(5) With θ^*_{ij} replacing θ_{ij}, calculate $S^2_{ij}(\theta_{ij})$, equation (4.13) and, hence, initial estimates, s^2_{ij}, of σ^2_{ij} from (4.15). In this last formula, use

G and H as the geometric and harmonic means [equations (4.3)] of $s_{ij}^2(\theta_{ij}^*)/(r_{ij} + 2)$.

(6) Calculate revised estimates of s_a^2, s_b^2, and s_c^2 using equations (2.5b-d).

(7) With these new estimates of σ_a^2, σ_b^2, and σ_c^2 and the estimates of σ_{ij}^2; resolve equations (3.1) except that σ^2 is replaced by the estimate of σ_{ij}^2 where σ^2 divides θ_{ij}^* and x_{ij}.

(8) Repeat (4) using the new estimates for θ_{ij}.

(9) Repeat (5).

(10) Repeat (6).

Repeat (7)-(10) until the results converge.

Notice that in the final solution of (3.1)--stage (7)--the matrix whose inverse is effectively obtained is the dispersion matrix of the (θ_{ij}^*) and should be made available.

APPENDIX 6: A Numerical Example

In this appendix, we describe the results of analyzing a simple case using the methods developed in the paper. Richmers and Todd (1967) give the following data, in their Table (8.21), taken from an experiment on the breaking strength of three fabrics at four temperatures with two replicates at each of the twelve combinations. We, therefore, have the case of constant

Fabric	Temperature 210	215	220	225
A	1.8	2.0	4.6	7.5
	2.1	2.1	5.0	7.9
B	2.2	4.2	5.4	9.8
	2.4	4.0	5.6	9.2
C	2.8	4.4	8.7	13.2
	3.2	4.8	8.4	13.0

numbers of replicates, and we assume that σ_{ij}^2 is also fixed but unknown at σ^2. We, therefore, have the simpler situation discussed in the bulk of the paper. The prior distribution suggested therein seems appropriate except that exchangeability of the column values (temperatures) ignores the fact that they are in sequence. But such information on ordering is neglected in the usual analysis of variance technique, so we do the same for comparison purposes. In the standard method, the 3 degrees of freedom associated with temperature would be broken up into linear and perhaps, quadratic terms: a parallel Bayesian analysis could easily be developed.

We took $\nu = 0$, $\nu_t = \nu_t = 1$ ($t = a, b, c$) in equations (2.5). These correspond to weak prior knowledge without causing convergence problems. (Values $\nu_t = 3$ were also tried with only a small effect on the results.)

The next table gives for each of the 12 cells the estimate θ^*_{ij} of the cell mean obtained from equation (7) with estimates from (2.5) of the variance components replacing the σ's. Also, included in brackets is the mean of the two original readings for that cell for comparison purposes. For each row and column there are similarly given the estimates α^*_i and β^*_j from (3.2) and (3.3) together with the data means in brackets for comparison.

Fabric	Temperature 210	215	220	225	
A	1.39 (1.95)	2.41 (2.05)	5.11 (4.80)	8.80 (7.70)	4.31 (4.13)
B	2.24 (2.30)	3.49 (4.10)	6.02 (5.50)	9.85 (9.60)	5.38 (5.38)
C	3.63 (3.00)	4.85 (4.60)	7.72 (8.55)	11.62 (13.10)	7.10 (7.31)
	2.53 (2.42)	3.66 (3.58)	6.26 (6.28)	9.95 (10.13)	

The estimates of the variances are $s^2 = 0.495$, $s_a^2 = 0.991$, $s_b^2 = 5.591$, $s_c^2 = 0.098$. These show a large effect of temperature, a smaller effect of fabric, and a small interaction term. The estimates θ^*_{ij} are, therefore, dominated by the additive effect of the two factors. These, displayed in the

borders of the table, show the usual shift toward the overall mean. For example, the value of β_1^*, the mean breaking strength at 210 is 2.53, greater than the observed mean of 2.42. The shift with the cell means is greater because of the almost complete removal of the interaction component. Thus, fabric A at 225 is estimated at 8.80 against an observed value of 7.70 which is a shift <u>away</u> from the mean. Notice that as a result of these shifts, the estimate of residual variance is at 0.495, much larger than the conventional value of 0.056 obtained from the 12 within-cell differences.

I am most grateful to David Christ and Gerald Isaacs who wrote the computer program and ran the above example. Their enthusiasm and expertise was most helpful and provided an illuminating insight into the merits of interactive computing.

References

Lindley, D. V. "The estimation of many parameters." Foundations of Statistical Inference. V. P. Godambe and D. A. Sprott (Eds.). Toronto: Holt, Rinehart, and Winston, 1971, 435-455.

Lindley, D. V., & Smith, A. F. M. "Bayes estimates for the linear model." Journal of the Royal Statistical Society, Series B, 1972, 34, 1-41.

Richmers, A. D., & Todd, H. N. Statistics: An Introduction. McGraw-Hill, 1967.

Snedecor, G. W. Statistical Methods. Ames: Iowa State College Press, 1956.